



COURSE REFERENCE

MiC Quality Online Courses

FREE ONLINE MODULE

“INTRODUCTION TO STATISTICS”

http://www.micquality.com/introductory_statistics/

www.micquality.com

STATS

PRIMER IN STATISTICS

PRIMER IN STATISTICS



This publication belongs to a series of **Course Reference** booklets that accompany our online courses. This booklet summarizes material covered in our **Primer in Statistics** online course.

For information about all our courses and **free Six Sigma resources** visit our Web site at www.micquality.com:

- σ [Free Online Module](#) "Introduction to Statistics"
- σ [Free Glossary](#) with 440+ Six Sigma and statistical terms
- σ [Free Online Excel Primer](#)
- σ [Reference tables](#), [Six Sigma and Normal Distribution calculators](#)

Glen Netherwood, MiC Quality

CONTENTS

Useful Statistical Measures	1
Normal Distribution	2
Useful Values from the Normal Distribution	3
Six Sigma 'Sigma Values'	3
Flow Charts.....	4
Check Sheets.....	5
Cause and Effect Diagrams	5
Histograms and Pareto Charts	6
Stem and Leaf Plots.....	7
Run Charts.....	8
Scatter Graphs	9
Correlation	9
Cusum Charts	10
Interactions.....	11
Multi-Vari Charts	12
Box Plots	13
Grouped Data	14
Percentiles	14
Control Charts.....	15
Range and Standard Deviation	16
Normal Probability Plots	17
Sampling Methods	18
Important Statistical Concepts.....	19
Measurement Scales.....	20

APPENDICES

Appendix 1 - Values of d_2 and d_2^*

Appendix 2 - Normal Distribution Tables

Appendix 3 - Limits on Runs for Run Charts

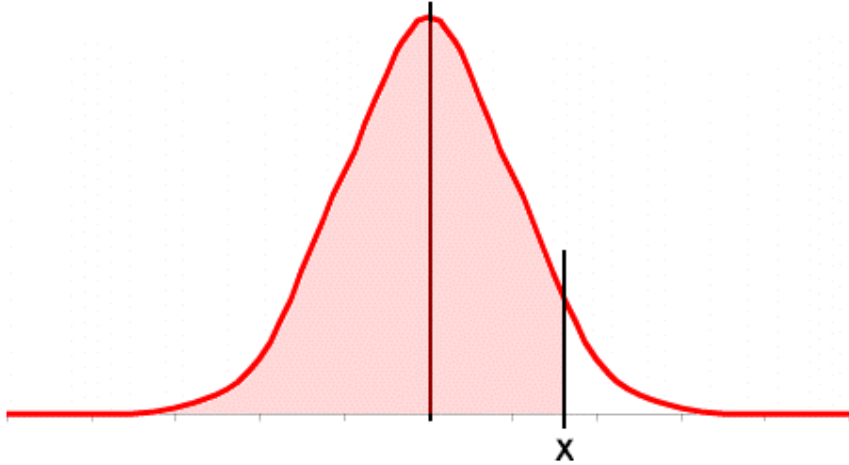
OUR STUDENTS SAY: **Jennifer McClare, Engineer, Canada**

"Very practical, lots of examples, easy to understand. Rather than just a review of math, the course was very applied with a number of very practical real-world examples. It showed me that I already knew enough to be making improvements in processes, but just didn't know how to apply it. The email support was very thorough and contained personal responses, not "canned" answers; individual attention was at least as good as in a classroom setting, if not better."

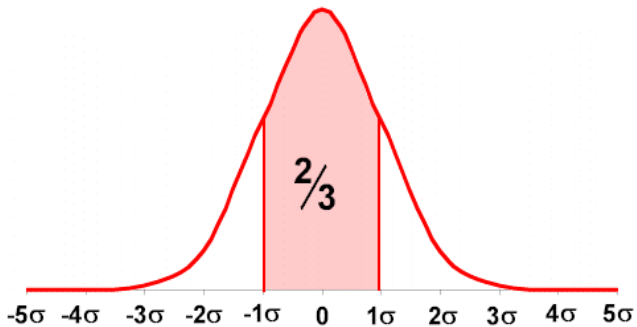
USEFUL STATISTICAL MEASURES

MEASURE	DESCRIPTION	FORMULA	EXCEL FUNCTION
Mean/ Average	The sum of the sample values divided by the sample size. The sample statistic \bar{x} is an unbiased estimate of the population mean μ .	$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$	=AVERAGE(data_range)
Median	The middle value after the sample values have been sorted into order by magnitude. If there are an even number of values in the sample, the average of the two middle values.		=MEDIAN(data_range)
Mode	The most common value in the sample.		=MODE(data_range)
Range	The difference between the largest and smallest values in the sample.		=MAX(data_range) - MIN(data_range)
Variance	An estimate of the variation or dispersion of the process from which the sample was drawn. The sample statistic ' s^2 ' is an unbiased estimator of the population parameter σ^2 .	$s^2 = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1}$	=VAR(data_range)
Standard Deviation	The square root of the variance. Often preferred as a measure of process variation. The sample statistic ' s ' is an estimator of the population parameter ' σ '. This method of calculating the standard deviation is known as the Root Mean Square Error (RMSE) method.	$s = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1}}$	=STDEV(data_range)

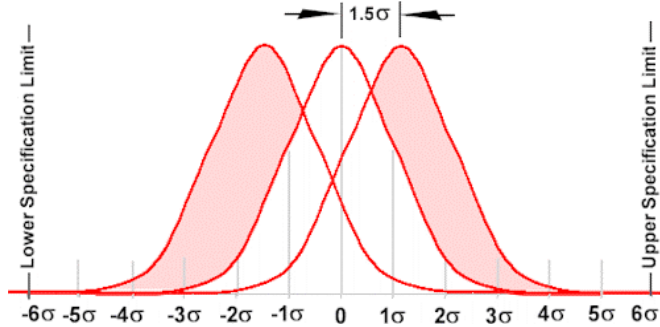
NORMAL DISTRIBUTION

DESCRIPTION	<p>The output of many types of processes can be represented by the normal distribution. The normal distribution is often used to estimate the proportion of process output that will lie within a specific range of values (for example, the proportion of the process output that will be within specification).</p> <p>The normal distribution shape results from the 'common cause' variation in a process. It can be used to reveal the presence of 'special causes'.</p>
DISTRIBUTION	
EXCEL FUNCTION	<p>The proportion of the process output smaller than 'X' can be found in Excel using:</p> <p>=NORMDIST (X, Mean, Standard Deviation, TRUE)</p>
TABLES	<p>The proportion of the process output smaller than 'X' can be also found from Normal Distribution Tables (see Appendix 2).</p>
Z-SCORE FORMULA	<p>The z-score is calculated from:</p> $Z = \frac{X - \mu}{\sigma}$ <p>'μ' process mean, estimated from the sample mean '\bar{x}'</p> <p>'σ' process standard deviation, estimated from the sample standard deviation 's'</p>
EXAMPLE	<p>A process produces bars with a mean length of 50mm and a standard deviation of 5, what proportion of the bars will be shorter than 64mm?</p> <p>Answer:</p> <p>The z-score is 2.80. From tables find that 0.9974 (99.74%) of the bars will be shorter than 64mm.</p>


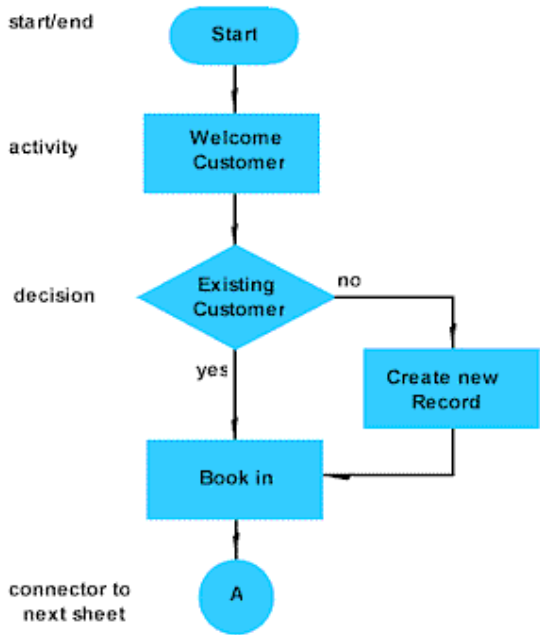
USEFUL VALUES FROM THE NORMAL DISTRIBUTION

DISTRIBUTION	About two thirds of the process output lie within one standard deviation either side of the process mean. Other values to remember:			
				
VALUES	Standard Deviations either side of the mean	Approximate amount	Exact amount	
	±1	two-thirds	68.27%	
	±2	95%	95.45%	
	±3	99.7%	99.73%	

SIX SIGMA 'SIGMA VALUES'

SIX SIGMA	The 'six sigma' approach uses the 'sigma value' to measure the number of DPMO (Defects per Million Opportunities). The calculation assumes that the process mean is 1.5 standard deviations from the target.			
				
VALUES (ppm – parts per million)	Sigma Level	Yield (%)	Defective ppm	
	1	30.23	697700	
	2	69.13	308700	
	3	93.32	66810	
	4	99.3790	6210	
	5	99.97670	233	
	6	99.999660	3.4	

FLOW CHARTS

CHART TYPE	Flow Chart
PROCESS PERSPECTIVE	<p>A process is defined as a series of activities aimed at the conversion of inputs into value added outputs</p>  <p>Activities in organizations, whether administrative transactions activities or production activities can be modeled as 'processes'.</p>
FLOW CHART	<p>A chart that shows the steps in a process. Preparing a flow chart should be one of the first steps in a process improvement activity.</p> <p>The flow chart should be based on close observation of what actually happens, and should not omit any steps.</p>
EXAMPLE	 <pre> graph TD Start([Start]) --> Welcome[Welcome Customer] Welcome --> Existing{Existing Customer} Existing -- no --> Create[Create new Record] Existing -- yes --> Book[Book in] Create --> Book Book --> A((A)) </pre>

CHECK SHEETS & CAUSE AND EFFECT DIAGRAMS

Check Sheet

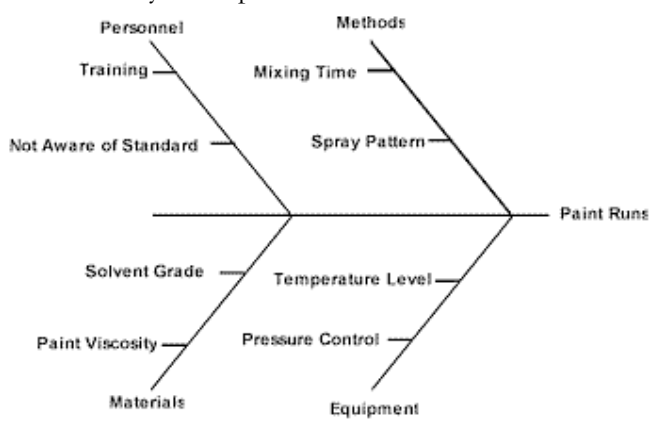
DESCRIPTION

A method for collecting data from a process; for example, data on the frequency and types of defects.

EXAMPLE

The example shows a check sheet used to record types of paint defects on an automotive part.

			Quantity		Pinholes	Dirt	Fibre	Clearcoat	Run	Solvent Boil	Mottled	Fried	Fisheye
Paint Date	Shift Booth	Color/Spoiler Type	Accept	Reject									
20/7	D2	XR Silver	9	3	/	/			/				
20/7	D3	XR White	10	2		//							
20/7	A2	DF Blue	14	1		/							
20/7	D1	DF Silver	11	4		/			///				
21/7	A3	JR Red	9	3	/	/			/				
22/7	D1	DF Blue	14	2		//							
23/7	A2	XR White	12	3		//			/				

CHART TYPE	Cause and Effect Diagrams
DESCRIPTION	Also called Ishikawa or Fishbone diagrams. A method for categorizing possible causes of a problem.
EXAMPLE	<p>The example shows only a few possible causes to illustrate the concept:</p> 

HISTOGRAMS AND PARETO CHARTS

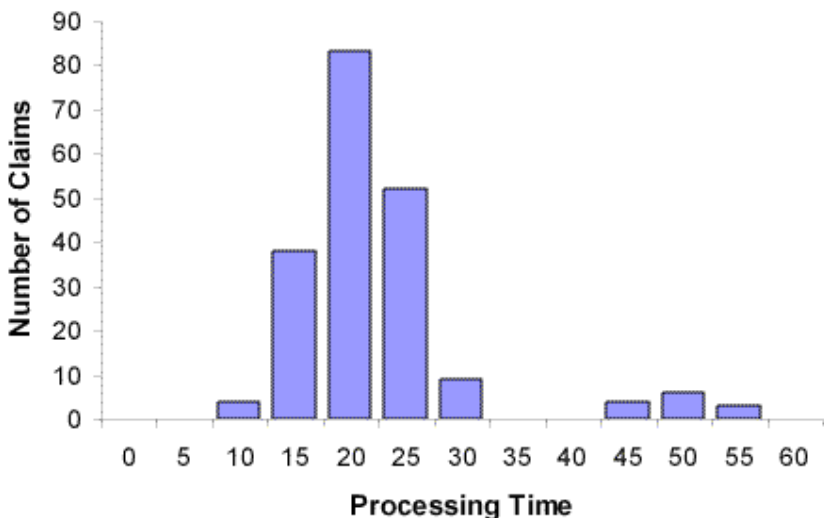
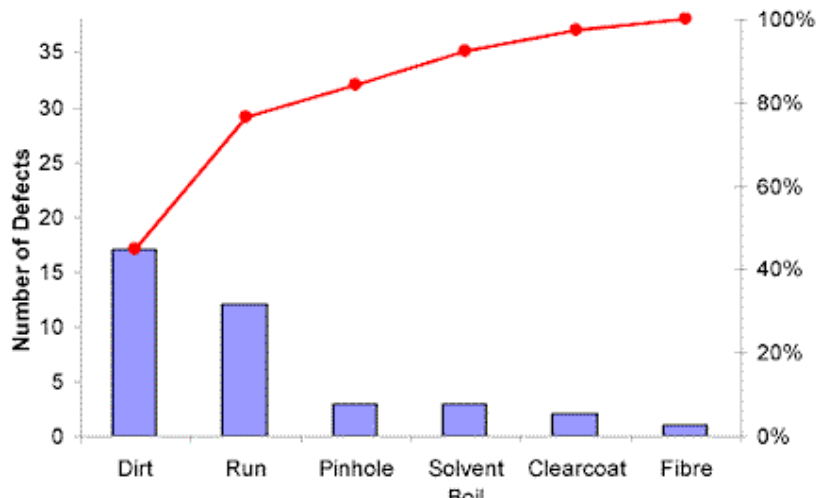
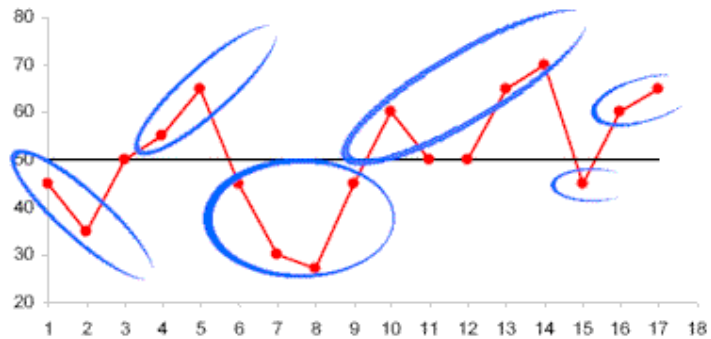
CHART TYPE	Frequency Histogram																		
DESCRIPTION	A useful way to represent the distribution of a set of data.																		
EXAMPLE	<p>The example shows processing time for insurance claims.</p>  <table border="1"> <caption>Data for Frequency Histogram Example</caption> <thead> <tr> <th>Processing Time (min)</th> <th>Number of Claims</th> </tr> </thead> <tbody> <tr><td>10-15</td><td>5</td></tr> <tr><td>15-20</td><td>38</td></tr> <tr><td>20-25</td><td>85</td></tr> <tr><td>25-30</td><td>52</td></tr> <tr><td>30-35</td><td>10</td></tr> <tr><td>45-50</td><td>5</td></tr> <tr><td>50-55</td><td>7</td></tr> <tr><td>55-60</td><td>4</td></tr> </tbody> </table>	Processing Time (min)	Number of Claims	10-15	5	15-20	38	20-25	85	25-30	52	30-35	10	45-50	5	50-55	7	55-60	4
Processing Time (min)	Number of Claims																		
10-15	5																		
15-20	38																		
20-25	85																		
25-30	52																		
30-35	10																		
45-50	5																		
50-55	7																		
55-60	4																		

CHART TYPE	Pareto Chart																					
DESCRIPTION	Used to sort the 'significant few' from the 'trivial many'. Often combines a sorted histogram and a cumulative graph.																					
CHART EXAMPLE	<p>The example shows the frequency of occurrence of types of paint defects on an automotive part..</p>  <table><thead><tr><th>Defect Type</th><th>Number of Defects</th><th>Cumulative Percentage (%)</th></tr></thead><tbody><tr><td>Dirt</td><td>17</td><td>40.9</td></tr><tr><td>Run</td><td>12</td><td>62.9</td></tr><tr><td>Pinhole</td><td>3</td><td>70.0</td></tr><tr><td>Solvent Boil</td><td>3</td><td>77.1</td></tr><tr><td>Clearcoat</td><td>2</td><td>82.9</td></tr><tr><td>Fibre</td><td>1</td><td>85.5</td></tr></tbody></table>	Defect Type	Number of Defects	Cumulative Percentage (%)	Dirt	17	40.9	Run	12	62.9	Pinhole	3	70.0	Solvent Boil	3	77.1	Clearcoat	2	82.9	Fibre	1	85.5
Defect Type	Number of Defects	Cumulative Percentage (%)																				
Dirt	17	40.9																				
Run	12	62.9																				
Pinhole	3	70.0																				
Solvent Boil	3	77.1																				
Clearcoat	2	82.9																				
Fibre	1	85.5																				

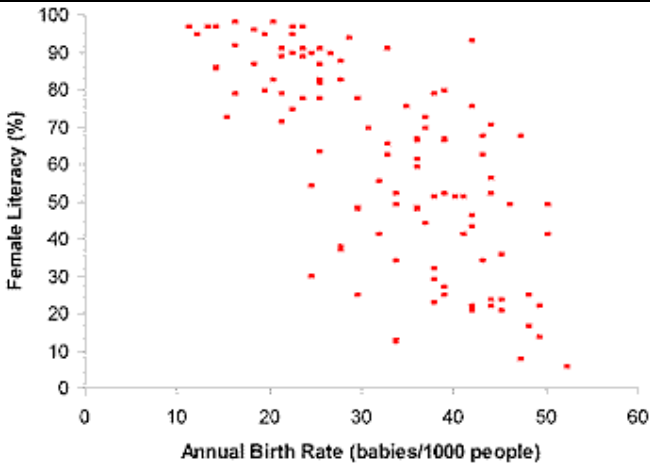
STEM AND LEAF PLOTS

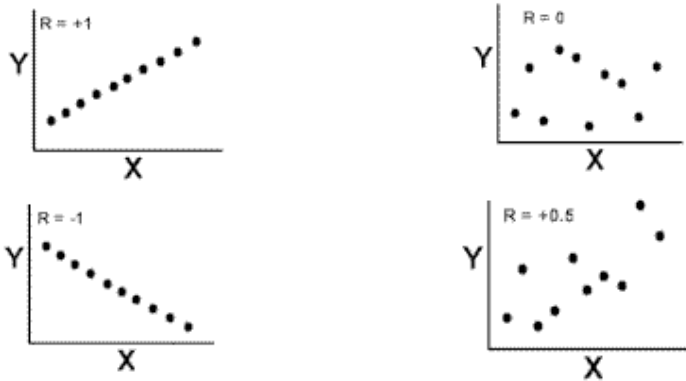
CHART TYPE	Stem and Leaf Plots																																								
DESCRIPTION	Stem and Leaf Plots are similar to histograms, but retain some, or all, of the original data values.																																								
EXAMPLE	The following values:																																								
	<table><tr><td>38</td><td>10</td><td>60</td><td>90</td><td>88</td></tr><tr><td>96</td><td>1</td><td>41</td><td>86</td><td>14</td></tr><tr><td>25</td><td>5</td><td>3</td><td>16</td><td>22</td></tr><tr><td>2</td><td>29</td><td>34</td><td>55</td><td>36</td></tr><tr><td>37</td><td>36</td><td>91</td><td>47</td><td>43</td></tr></table>					38	10	60	90	88	96	1	41	86	14	25	5	3	16	22	2	29	34	55	36	37	36	91	47	43											
	38	10	60	90	88																																				
	96	1	41	86	14																																				
	25	5	3	16	22																																				
	2	29	34	55	36																																				
	37	36	91	47	43																																				
	Would form a stem and leaf plot:																																								
	<table><tr><th>Frequency</th><th>Stem</th><th>Leaves</th></tr><tr><td>5</td><td>0</td><td>1235</td></tr><tr><td>3</td><td>1</td><td>046</td></tr><tr><td>3</td><td>2</td><td>259</td></tr><tr><td>5</td><td>3</td><td>46678</td></tr><tr><td>3</td><td>4</td><td>137</td></tr><tr><td>1</td><td>5</td><td>5</td></tr><tr><td>1</td><td>6</td><td>0</td></tr><tr><td>0</td><td>7</td><td></td></tr><tr><td>2</td><td>8</td><td>68</td></tr><tr><td>3</td><td>9</td><td>116</td></tr><tr><td>0</td><td>10</td><td></td></tr></table>					Frequency	Stem	Leaves	5	0	1235	3	1	046	3	2	259	5	3	46678	3	4	137	1	5	5	1	6	0	0	7		2	8	68	3	9	116	0	10	
	Frequency	Stem	Leaves																																						
5	0	1235																																							
3	1	046																																							
3	2	259																																							
5	3	46678																																							
3	4	137																																							
1	5	5																																							
1	6	0																																							
0	7																																								
2	8	68																																							
3	9	116																																							
0	10																																								
There are two values with a stem of '8', the values 86 and 88. These go into the row with stem '8'; the leaves are '6' and '8'.																																									

RUN CHARTS

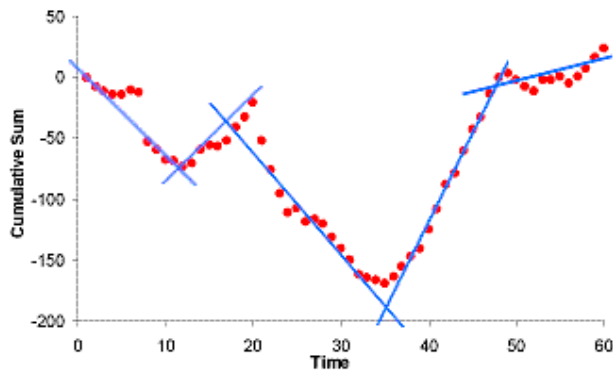
CHART TYPE	Run Charts								
DESCRIPTION	<p>Charts that plot data in time sequence and so identify patterns that suggest 'special causes'.</p> <p>The center line of a run chart is the median. A run is one or more consecutive points on the same side of the median. The run stops when it crosses the median. Points on the median do not stop a run, but are not counted in the run.</p>								
EXAMPLE									
TESTS	<p>Test 1: The number of runs The number of runs is checked against the table in Appendix 3.</p> <p>Test 2: Too many points in a run</p> <ul style="list-style-type: none"> less than 20 useful observations, 7 or more points 20 or more useful observations, 8 or more points <p>"Useful observations" exclude points on the median line.</p> <p>Test 3: Trend An unusually long sequence of consecutive points either ascending or descending:</p> <table border="1"> <thead> <tr> <th>Total Number of Points</th><th>Consecutive Ascending or Descending Points</th></tr> </thead> <tbody> <tr> <td>5 to 8</td><td>5 or more</td></tr> <tr> <td>9 to 20</td><td>6 or more</td></tr> <tr> <td>more than 20</td><td>7 or more</td></tr> </tbody> </table> <p>Count points on the median, but ignore points that repeat the preceding value.</p> <p>Test 4: Zig-Zag Fourteen or more consecutive points alternatively up and down.</p>	Total Number of Points	Consecutive Ascending or Descending Points	5 to 8	5 or more	9 to 20	6 or more	more than 20	7 or more
Total Number of Points	Consecutive Ascending or Descending Points								
5 to 8	5 or more								
9 to 20	6 or more								
more than 20	7 or more								

SCATTER GRAPHS & CORRELATION

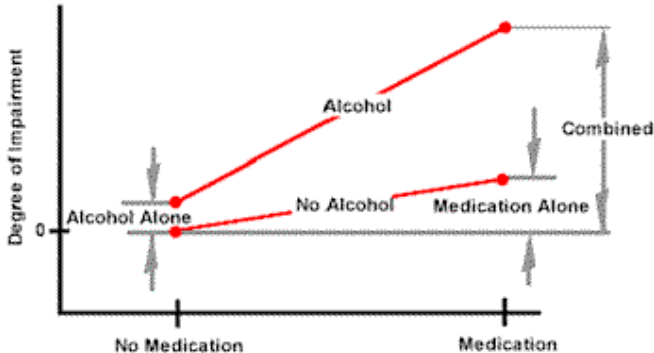
CHART TYPE	Scatter Graphs
DESCRIPTION	Scatter graph are used to explore associations between two variables. They are normally used when the data form natural pairs, and where there are many pairs. Standard X-Y graphs are normally used to explore a causal relationship between X and Y; typically X is varied systematically and the effect on Y measured.
EXAMPLE	
COMMENTS	Female Literacy and Birth Rate are associated, but there is not necessarily a causal relationship.

Correlation	
DESCRIPTION	Correlation is a measure of the strength of the relationship between the input and the output of a process. Correlation is measured by the 'Pearson Product Moment Correlation', known as 'R'. The value of 'R' varies from +1 to -1.
EXAMPLES	
COMMENTS	An R value of + 1 is perfect correlation. Values between -0.5 and + 0.5 show weak relationships.

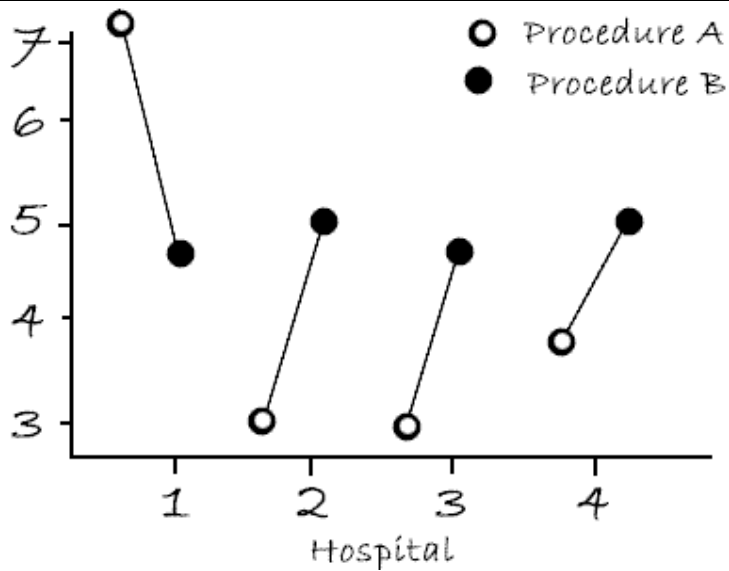
CUSUM CHARTS

CHART TYPE	Cusum Charts																									
DESCRIPTION	<p>Cusum charts plot the cumulative deviation from a specified value. The example uses a specified value of 190:</p> <table><tr><th></th><th>Data</th><th>Difference</th><th>Calculations</th><th>Cumulative Sum</th></tr><tr><td>1</td><td></td><td></td><td>0</td><td></td></tr><tr><td>2</td><td>189</td><td>-1</td><td>0 + -1 = -1</td><td>-1</td></tr><tr><td>3</td><td>183</td><td>-7</td><td>-1 + -7 = -8</td><td>-8</td></tr><tr><td>4</td><td>192</td><td>+2</td><td>-8 + 2 = -6</td><td>-6</td></tr></table> <p>The advantage of the cusum chart is that is very sensitive to long-term, small changes to the process mean.</p>		Data	Difference	Calculations	Cumulative Sum	1			0		2	189	-1	0 + -1 = -1	-1	3	183	-7	-1 + -7 = -8	-8	4	192	+2	-8 + 2 = -6	-6
	Data	Difference	Calculations	Cumulative Sum																						
1			0																							
2	189	-1	0 + -1 = -1	-1																						
3	183	-7	-1 + -7 = -8	-8																						
4	192	+2	-8 + 2 = -6	-6																						
EXAMPLE																										
COMMENTS	<p>The intersections of the lines represent the turning points. Thus a ‘special cause’ has affected the process at each of the four turning points in the example chart.</p> <p>For long-term monitoring of processes a variation on the basic cusum chart known as the ‘tabular cusum chart’ is preferable. The tabular cusum chart allows for the small variations in the process mean that are acceptable and unavoidable in most processes.</p>																									

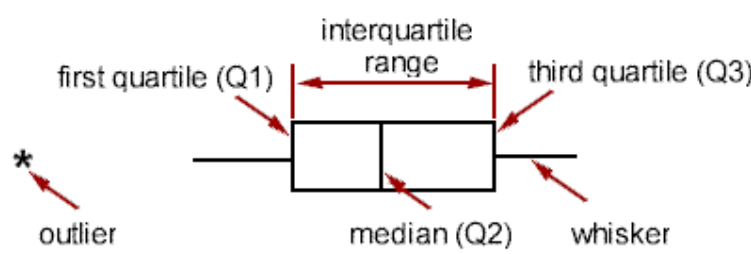
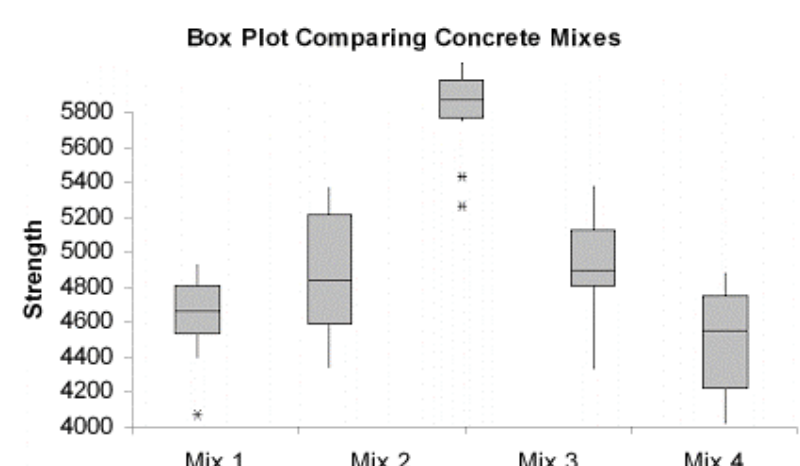
INTERACTIONS

Interactions	
DESCRIPTION	Where a process has several inputs, interactions may exist between the inputs. This means that the effect of one input depends on the value of one or more of the other inputs.
EXAMPLE	
COMMENTS	<p>Some medications interact with alcohol. If a person has either the medication or a moderate amount of alcohol their judgment is not significantly affected. If they combine the two, the effect is substantial.</p> <p>The Design of Experiments (DOE) is the most effective method for analyzing processes that have multiple inputs, particularly if there may be interactions between the inputs.</p>

MULTI-VARI CHARTS

CHART TYPE	Multi-Vari Charts																											
DESCRIPTION	<p>Multi-vari charts are a graphical method used to show the effect of more than one input variable on the output, including any interaction between the input variables.</p> <p>This is illustrated by an example. The table shows the Average Length of Stay (LOS) for two procedures, 'A' and 'B' at four different hospitals.</p> <table><thead><tr><th>Hospital</th><th>Procedure</th><th>LOS</th></tr></thead><tbody><tr><td>1</td><td>A</td><td>7.25</td></tr><tr><td>1</td><td>B</td><td>4.75</td></tr><tr><td>2</td><td>A</td><td>3.00</td></tr><tr><td>2</td><td>B</td><td>5.00</td></tr><tr><td>3</td><td>A</td><td>3.00</td></tr><tr><td>3</td><td>B</td><td>4.75</td></tr><tr><td>4</td><td>A</td><td>3.75</td></tr><tr><td>4</td><td>B</td><td>5.00</td></tr></tbody></table>	Hospital	Procedure	LOS	1	A	7.25	1	B	4.75	2	A	3.00	2	B	5.00	3	A	3.00	3	B	4.75	4	A	3.75	4	B	5.00
Hospital	Procedure	LOS																										
1	A	7.25																										
1	B	4.75																										
2	A	3.00																										
2	B	5.00																										
3	A	3.00																										
3	B	4.75																										
4	A	3.75																										
4	B	5.00																										
EXAMPLE																												
COMMENTS	<p>The chart clearly shows that Procedure A at Hospital 1 is the odd one out. Multi-vari charts can show the effects of more than two input variables. If there were three input variables, there would be a separate chart for each level of the third variable.</p>																											

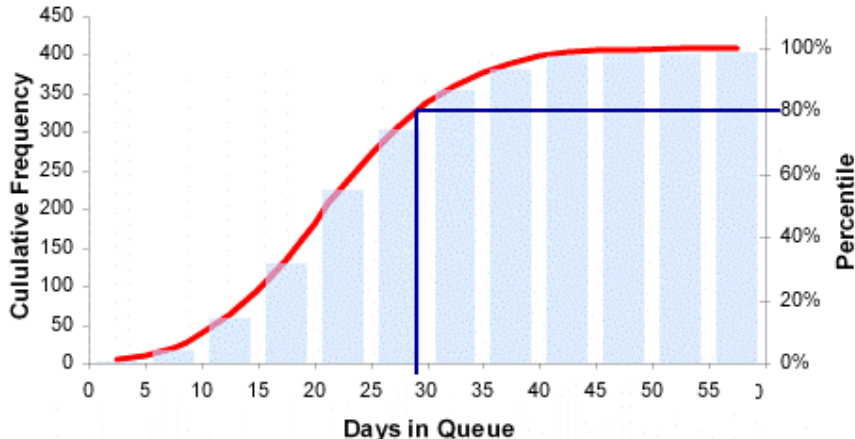
Box Plots

CHART TYPE	Box Plot
DESCRIPTION	<p>A graphical method for representing a data set.</p>  <ul style="list-style-type: none"> • outliers are smaller than $Q1 - 1.5 \times (Q3 - Q1)$ or greater than $Q3 + 1.5 \times (Q3 - Q1)$ • whiskers are the largest and smallest data values that are not outliers
EXAMPLE	<p>The example shows a box plot used to compare the strength of five mixes of concrete:</p> 

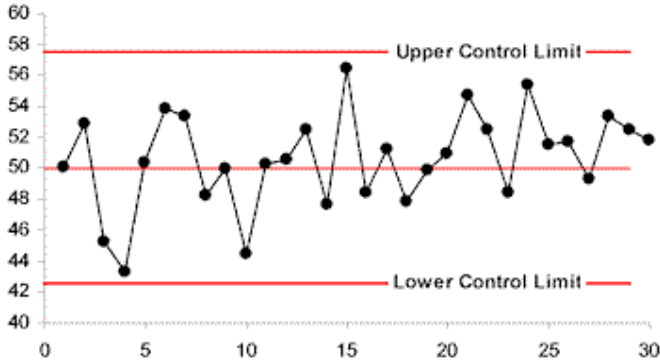
GROUPED DATA

MEASURE	DESCRIPTION	FORMULA
Mean of Grouped Data	Used where data are collected in ranges of values, rather than individual values (e.g., census data). The number of individual data values would typically be large.	$\bar{x} = \frac{\sum_{j=1}^k f_j M_j}{n}$ <p> f_j - frequency for group 'j' M_j - midpoint of the range (smallest + largest)/2 n - the total number of samples k - the number of groups </p>
Standard Deviation of Grouped Data		$s = \sqrt{\frac{\sum_{j=1}^k f_j (M_j - \bar{x})^2}{n-1}}$

PERCENTILES

DESCRIPTION	Percentiles show the cumulative frequency and are often associated with grouped data.
EXAMPLE	<p>The chart shows how a frequency 'ogive' is created from grouped data (displayed as a cumulative histogram). It also shows, as an example, how the number of days corresponding to the 80th percentile is obtained.</p> 

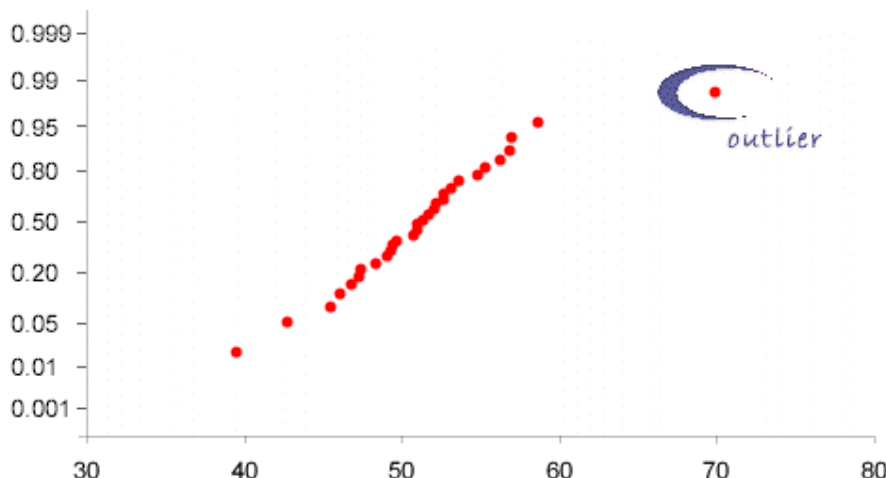
CONTROL CHARTS

CHART TYPE	Control Charts
DESCRIPTION	<p>Control charts are used to monitor processes. They are introduced after a state of 'statistical control' has been achieved. The purpose of control charts is to achieve timely identification of 'special causes' entering the process. There are many types of control charts, and control charts are used to monitor both variables and attributes. For comprehensive coverage see the MiC Quality Statistical Process Control and Advanced SPC courses at www.micquality.com</p>
EXAMPLE	
COMMENTS	<p>The control limits are usually located at three standard deviations from the mean. Three standard deviations are selected because it is a reasonable compromise between the probability of a 'Type I Error' and a 'Type II Error'. In healthcare some experts advocate using two standard deviations.</p>
WESTERN ELECTRIC RULES	<p>Various patterns are used to signify a 'special cause'. The 'Western Electric' rules are often used.</p> <ul style="list-style-type: none"> Rule 1: A single point outside the three sigma control limits, on either side. Rule 2: Any two of the last three points outside the two sigma line, on the same side Rule 3: Any four of the last five points outside the one sigma line, on the same side Rule 4: Eight consecutive points on the same side of the center line Rule 5: Six consecutive points trending either up or down Rule 6: Fourteen consecutive points alternating up and down
COMMENTS	<p>In addition about two thirds of the points should be within the one standard deviation of the mean, and about 95% within 2 standard deviations.</p>

RANGE AND STANDARD DEVIATION

DESCRIPTION	The process standard deviation can be estimated by calculating the average range of a number of samples, each of size ‘n’, and dividing by a constant.																														
n >= 15	<p>If the number of samples is greater than about 15:</p> $s = \frac{\bar{R}}{d_2}$ <p>Values of d₂:</p> <table><tr><td>n</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td></tr><tr><td>d₂</td><td>1.128</td><td>1.693</td><td>2.059</td><td>2.326</td><td>2.534</td><td>2.704</td><td>2.847</td><td>2.970</td><td>3.078</td><td>3.173</td><td>3.259</td><td>3.336</td><td>3.407</td><td>3.472</td></tr></table>	n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	d ₂	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078	3.173	3.259	3.336	3.407	3.472
n	2	3	4	5	6	7	8	9	10	11	12	13	14	15																	
d ₂	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078	3.173	3.259	3.336	3.407	3.472																	
n < 15	<p>If the number of samples is less than 15:</p> $s = \frac{\bar{R}}{d_2^*}$ <p>Values of d₂[*] are tabulated in Appendix 1.</p>																														
NOTE	d₂[*] is mainly used in Measurement Systems Analysis/ Gage R&R Studies where the number of samples is unavoidably small. In other applications at least 30 samples are normally used for reliable results.																														
EXAMPLE	<p>Thirty samples, each containing five items are taken from a process. The average range of the thirty samples is 5 mm. Estimate the standard deviation of the process.</p> <p>Answer: The value of d₂ is 2.326, the standard deviation is 2.15</p>																														

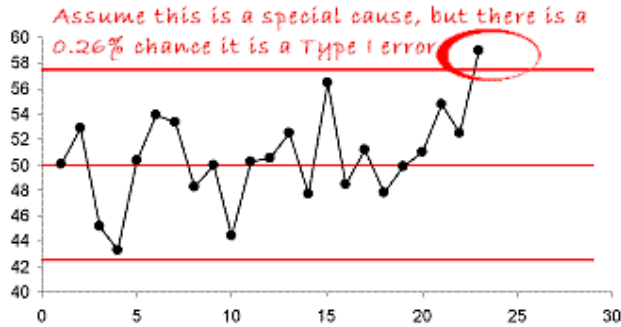
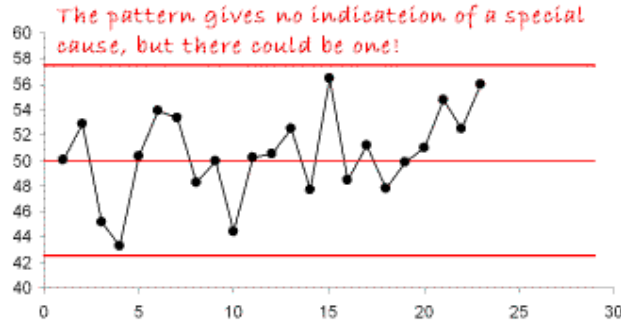
NORMAL PROBABILITY PLOTS

CHART TYPE	Normal Probability Plot								
DESCRIPTION	A type of graph used to check if a sample conforms to a normal distribution, or to identify values that do not conform to a normal distribution (outliers).								
EXAMPLE	<div><p>The 'y' scale of the normal probability plot is non-linear. It can be created using special graph paper or can be created in Excel using the function '=NORMSINV' as a transform as follows:</p><table><tr><th>Index</th><th>Fi</th><th>Y</th><th>Ordered Data (x)</th></tr><tr><td>i</td><td>$F_i = \frac{i-0.5}{n}$</td><td>=NORMSINV(F_i)</td><td>x_i</td></tr></table><p>Index values run from 1 through n, where n is the number of values in the sample.</p></div>	Index	Fi	Y	Ordered Data (x)	i	$F_i = \frac{i-0.5}{n}$	=NORMSINV(F _i)	x _i
Index	Fi	Y	Ordered Data (x)						
i	$F_i = \frac{i-0.5}{n}$	=NORMSINV(F _i)	x _i						

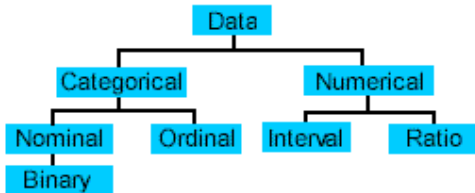
SAMPLING METHODS

Sampling Types	
Sampling	Inferential statistics is used to draw conclusions about a population based on a relatively small sample. The conclusions are only meaningful if the sample is representative of the population.
Simple Random Sampling	In a simple random sample, the sampling is carried out in a way that ensures that every member of the population has an equal chance of being selected. This can be done by numbering each item in the population, and then picking numbers from a hat, as in a lottery. An easier approach is to use a computer to generate random numbers.
Stratified Sampling	<p>This involves splitting the population into categories and then taking a random sample from each category. The size of the samples are proportional to the size of the category.</p> <p>Suppose a company wants to carry out a survey of employee satisfaction. A simple random sample would select employees at random. A stratified sample might select employees proportionally from each department, and level of management. In a small mixed gender group it may be appropriate to ensure that males and females are proportionally represented.</p>
Convenience Sampling	<p>A simple random sample is often impractical because some items are difficult to access, for example in products that are palletized.</p> <p>Sampling from the easy to access items may be acceptable if there is evidence that they are representative of the remainder of the batch.</p>
Sample Homogeneity	In some types of test the sample is taken from an area that may not be homogeneous, for example a 500 gram sample of soil may represent 10 acres of land. In this case sub-samples are taken from over the area and thoroughly mixed to make sure it is representative of the plot.

IMPORTANT STATISTICAL CONCEPTS

Important Statistical Concepts	
Inferential and Descriptive Statistics	<p>Descriptive (enumerative) statistics are used to summarize and describe important features of the data, such as the centering, spread or normality.</p> <p>Inferential (analytical) statistics uses information obtained from a sample to draw conclusions (make inferences) about the population as a whole.</p>
Common and Special Cause Variation	<p>Common cause variation refers to the many small factors, many of which will not have been identified, that contribute to process variation. If only common cause variation is present the process response will conform to a normal distribution.</p> <p>Special cause variation results from identifiable factors that have a significant effect on the process, larger than the common cause variation. Special causes can be identified and should, where possible, be eliminated.</p>
Statistical Control	<p>A process is in statistical control when only common cause variation is present, and when the statistical properties of the distribution do not vary with time.</p>
Type I and Type II Errors	<p>In the context of control charts, a Type I Error occurs when the patterns on the control chart indicate that the process is not in a state of statistical control, but in reality the process is in control:</p>  <p>A Type II Error occurs when the process is not in a state of statistical control, but no evidence of this has shown up on the control chart:</p> 

MEASUREMENT SCALES

Measurement Scales	
DESCRIPTION	When you collect sample data you use a 'Measurement Scale'. The choice of scale affects the amount of information you will get from the data, and the mathematical operations that you can use with it. There are four types of measurement scale, nominal, ordinal, interval and ratio.
Taxonomy	 <pre> graph TD Data[Data] --> Categorical[Categorical] Data --> Numerical[Numerical] Categorical --> Nominal[Nominal] Categorical --> Ordinal[Ordinal] Nominal --> Binary[Binary] Numerical --> Interval[Interval] Numerical --> Ratio[Ratio] </pre>
Nominal (Attribute)	<p>This is the most basic measurement scale. Data are placed into categories that cannot be sorted into a logical order. An example would be marital status: single, married, divorced. Nominal data are also known as 'attribute data'.</p> <p>arithmetic: <i>equal to</i> statistical: <i>mode</i></p>
Ordinal	<p>Data are sorted into categories. The categories can be placed into a logical order, but the intervals between the categories are undefined. Often used for order of preference.</p> <p>arithmetic: <i>comparison (equal to, greater than, less than)</i> statistical: <i>median, mode</i></p>
Interval	<p>Items are placed on a scale, with intervals that can be measured numerically. The numerical scale can be linear or non-linear (e.g. logarithmic). Zero does not mean the absence of the entity. Rating customer satisfaction on a scale of 1 to 10 would be an example of an interval scale. The Fahrenheit or Centigrade temperature scales are also interval scales because zero does not imply an absence of temperature.</p> <p>arithmetic: <i>comparison, addition, subtraction NOT multiplication, division.</i> statistical: <i>mean, median, variance</i></p>
Ratio	<p>Similar to an interval scale with the additional constraint that zero means the absence of the entity. Length or weights are measured on a ratio scale.</p> <p>arithmetic: <i>comparison addition, subtraction, multiplication, division</i> statistical: <i>mean, median, variance</i></p>

APPENDICES

APPENDIX 1: VALUES OF d_2 AND d_2^*

# Samples (k)	Size of Sample (n)													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.414	1.912	2.239	2.481	2.673	2.830	2.963	3.078	3.179	3.269	3.350	3.424	3.491	3.553
2	1.2879	1.805	2.151	2.405	2.604	2.768	2.906	3.025	3.129	3.221	3.305	3.380	3.449	3.513
3	1.231	1.769	2.120	2.379	2.581	2.747	2.886	3.006	3.112	3.205	3.289	3.366	3.435	3.499
4	1.206	1.750	2.105	2.366	2.570	2.736	2.877	2.997	3.103	3.197	3.282	3.358	3.428	3.492
5	1.191	1.739	2.096	2.358	2.563	2.730	2.871	2.992	3.098	3.192	3.277	3.354	3.424	3.488
6	1.181	1.731	2.090	2.353	2.558	2.726	2.867	2.988	3.095	3.189	3.274	3.351	3.421	3.486
7	1.173	1.726	2.085	2.349	2.555	2.723	2.864	2.986	3.092	3.187	3.272	3.349	3.419	3.484
8	1.168	1.721	2.082	2.346	2.552	2.720	2.862	2.984	3.090	3.185	3.270	3.347	3.417	3.482
9	1.164	1.718	2.080	2.344	2.550	2.719	2.860	2.982	3.089	3.184	3.269	3.346	3.416	3.481
10	1.160	1.716	2.077	2.342	2.549	2.717	2.859	2.981	3.088	3.183	3.268	3.345	3.415	3.480
11	1.157	1.714	2.076	2.340	2.547	2.716	2.858	2.980	3.087	3.182	3.267	3.344	3.415	3.479
12	1.155	1.712	2.074	2.3439	2.546	2.715	2.857	2.979	3.086	3.181	3.266	3.343	3.414	3.479
13	1.153	1.710	2.073	2.338	2.545	2.714	2.856	2.978	3.085	3.180	3.266	3.343	3.413	3.478
14	1.151	1.709	2.072	2.337	2.545	2.714	2.856	2.978	3.085	3.180	3.265	3.342	3.413	3.478
15	1.150	1.708	2.071	2.337	2.544	2.713	2.855	2.977	3.084	3.179	3.265	3.342	3.412	3.477
d_2	1.128	1.693	2.059	2.326	2.534	2.704	2.847	2.970	3.078	3.173	3.259	3.336	3.407	3.472

Duncan A. J (1986), Quality Control and Industrial Statistics Appendix D3

APPENDIX 2: NORMAL DISTRIBUTION TABLES



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.40	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.30	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.20	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.10	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

APPENDIX 2:

NORMAL DISTRIBUTION TABLES CONTINUED



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

APPENDIX 3: LIMITS ON RUNS FOR RUN CHARTS

Number of Points not on Median	Lower Limit	Upper Limit	Number of Points not on Median	Lower Limit	Upper Limit
10	3	8	34	12	23
11	3	9	35	13	23
12	3	10	36	13	24
13	4	10	37	13	25
14	4	11	38	14	25
15	4	12	39	14	26
16	5	12	40	15	26
17	5	13	41	16	26
18	6	13	42	16	27
19	6	14	43	17	27
20	6	14	44	17	28
21	7	15	45	17	29
22	7	16	46	17	30
23	8	16	47	18	30
24	8	17	48	18	31
25	9	17	49	19	31
26	9	18	50	19	32
27	9	19	60	24	37
28	10	19	70	28	43
29	10	20	80	33	48
30	11	20	90	37	54
31	11	21	100	42	59
32	11	22	110	46	65
33	11	22	120	51	70

MiC Quality Online Courses

SIX SIGMA PRIMER

This course provides an in-depth introduction to Six Sigma. It takes you through each stage of the DMAIC sequence using case studies to show you how, and when, to use the most important methods and tools.

TOPICS: Deployment, Six Sigma metrics, the DMAIC sequence, Lean methods, Design For Six Sigma (DFSS).

Recommended for everyone implementing Six Sigma or studying for a Black Belt.

STATISTICAL PROCESS CONTROL (SPC)

You will learn how to carry out process capability studies, use control charts effectively, and use the results to improve your processes.

TOPICS: Variation; process capability Cp, Cpk; process performance Pp, Ppk; X-Bar and R control charts; attribute control charts p, np, c, u.

A must for everyone involved in quality management, ISO9000, Six Sigma.

ADVANCED STATISTICAL PROCESS CONTROL

Covers 10 types of control charts for a variety of process situations, including short and high volume, as well as supporting process improvement and Six Sigma initiatives.

TOPICS: Given Standard, X-bar & s, Median, Demerits Per Unit (U), Individual and Moving Range (XmR), Moving Average & EWMA, CuSum; Short Run SPC; PRE-control.

Recommended for everyone involved in quality management, ISO9000 and Six Sigma.

PRIMER IN STATISTICS

An introduction to statistics and process improvement tools.

TOPICS: Mean, Median, Mode, Range; Histograms; Pareto Chart; Box Plots; Variance; Quartiles, Percentiles; Inferential Statistics; Normal Distribution; Range and Standard Deviation; Normal Probability Plots, Stem and Leaf Plots, Flow Charts, Process Improvement Tools

A must for everyone involved in quality management, ISO9000, Six Sigma.

ADVANCED STATISTICS

Comprehensive coverage of the statistical methods for engineers and scientists. An excellent preparation for the American Society for Quality Certified Six Sigma Black Belt (ASQ CSSBB) exam.

TOPICS: Confidence Intervals; t-distribution; Hypothesis Testing; t-tests; Type I and II errors and Power; Chi-Square Distribution; Contingency Tables; Regression Analysis; Correlation; ANOVA; Probability; Binomial, Poisson & Hypergeometric Distributions

Recommended for scientists, engineers, Six Sigma Black Belts and Master Black Belts.

DESIGN OF EXPERIMENTS (DOE)

A practical guide for people who need to improve their processes using experimental design.

TOPICS: Full and Fractional Factorial Designs; Design Resolution; Hypothesis Testing; ANOVA; Analysis of Residuals; Screening Designs; Plackett-Burman Designs

A must for Six Sigma Black Belts and Master Black Belts; recommended for engineers.

ADVANCED DESIGN OF EXPERIMENTS

The course shows how to analyze the advanced methods of experimental design using Minitab.

TOPICS: Taguchi Signal to Noise Ratio and Taguchi designs, Response Surface Designs, "Hill climbing" approach for process optimum, Mixture Designs

A must for Six Sigma Black Belts and Master Black Belts; recommended for engineers.

MEASUREMENT SYSTEMS ANALYSIS (MSA/GAGE R&R)

SPC and DOE rely on the integrity of the measurement systems. This course provides a thorough treatment on how to evaluate and improve measurement systems.

TOPICS: Control Chart Methods; Repeatability & Reproducibility; Gage R & R Studies; Evaluating the Results; Using Minitab; ANOVA Methods; Capability, Bias, Linearity & Stability; Attribute Studies - long and short methods.

Recommended for quality managers, Six Sigma Black Belts and Master Black Belts.

"The fundamentals of statistics have been explained in a beautiful manner which makes them easy to understand."

"I really appreciate your clear and practical explanations and the simulations! I've never really understood statistics very well. Your Primer was VERY helpful. I think I'm starting to understand this stuff!"

About MiC Quality

MiC Quality is a global provider of e-learning solutions. We provide online courses in the statistical methods used in quality assurance, process improvement, research and development, and **Six Sigma** programs. The courses are ideal for quality professionals, engineers, scientists, managers and supervisors who need to use statistics in their work. Our customers come from many industries including healthcare, manufacturing, biotechnology, electronics, IT, research and pharmaceuticals.

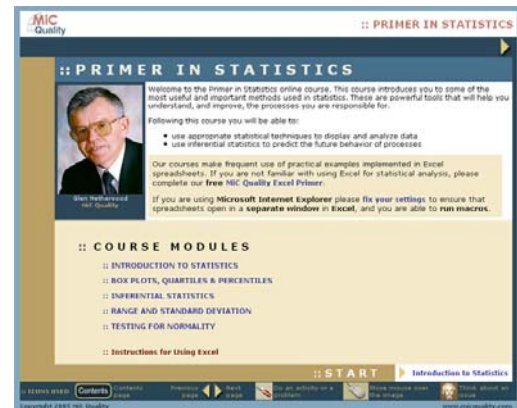
Benefits of MiC Quality online courses:

- :: **interactive** with exercises, simulations and case studies
- :: **extensive support** with individual coaching and feedback
- :: **comprehensive** with about 30 hours of in-depth study per course
- :: **flexible** self-paced learning available anywhere, anytime
- :: **effective** in developing practical knowledge and hands-on skills

MiC Quality online courses include:

- :: Six Sigma Primer
- :: Statistical Process Control (SPC)
- :: Advanced Statistical Process Control (SPC)
- :: Primer in Statistics
- :: Advanced Statistics
- :: Design of Experiments (DOE)
- :: Advanced Design of Experiments (DOE)
- :: Measurement Systems Analysis (MSA)/ Gage R&R

(see previous page for more information)



Solutions for Organizations and Individuals

Six Sigma, Process Improvement Programs

Use MiC Quality self-paced e-learning for the statistical component of Six Sigma methodology and reduce the need for expensive class-based courses.

Quality Management Systems

Provide training for group and individual staff members in statistical quality assurance methods to support **ISO9000**, **TS16949** and supplier accreditation requirements.

Professional Development for Associations

Form partnerships to provide members with professional development at discounted prices.

Career Development for Individuals

Support your professional career development, including the American Society for Quality qualifications of **Certified Quality Engineer (CQE)**, **Six Sigma Black Belt (SSBB)** and **Green Belt (SSGB)**.

Licenses and Partnerships

Site and group licenses are available. We welcome opportunities for new partnerships. Please **contact** us for more information.



Dragos Gabriel Marin

Analyst, Pratt & Whitney

"When I started the course my experience in statistics was a very traumatizing course at the university plus a number of unsuccessful

attempts of studying SPC from books. Now, at the end of the course, I can say that yes, **I understand the concepts, and I will apply them.** It is very well designed, the **e-mail support is excellent**, and it is **affordable.**"

- :: Try our **FREE** module
- :: Take our **FREE** Excel Primer
- :: **ENROLL** in our courses

Go online to:

www.micquality.com

Contact Information

MiC Quality

The Hill, 44 Sprys Lane,
PO Box 655, Hurstbridge
VIC 3099 Australia

Web: **www.micquality.com**

Freecall in the US: **1-888-240-5504**

Tel in Australia: **+61 3 9718 1112**

Fax in Australia: **+61 3 9718 1113**

Fax in the US: **(815) 846 8487**